Crystal structure of a dUTPase

Eila S. Cedergren-Zeppezauer*, Gunilla Larsson†, Per Olof Nyman†, Zbigniew Dauter‡ & Keith S. Wilson‡

* Department of Zoological Cell Biology, Wenner-Gren Institute, University of Stockholm, S-10691 Stockholm, Sweden
† Department of Biochemistry, Chemical Centre, University of Lund, S-22100 Lund, Sweden
‡ European Molecular Biology Laboratory (EMBL), c/o DESY, Notkestrasse 85, D-2000 Hamburg 52, Germany

THE enzyme dUTPase catalyses the hydrolysis of dUTP[1] and maintains a low intracellular concentration of dUTP so that uracil cannot be incorporated into DNA[2]. dUTPase from *Escherichia coli* is strictly specific for its dUTP substrate,[3] the active site discriminating between nucleotides with respect to the sugar moiety as well as the pyrimidine base. Here we report the three-dimensional structure of *E. coli* dUTPase determined by X-ray crystallography at a resolution of 1.9 Å. The enzyme is a symmetrical trimer, and of the 152 amino acid residues in the subunit, the first 136 are visible in the crystal structure. The tertiary structure resembles a jelly-roll fold and does not show the 'classical' nucleotide-binding domain. In the quaternary structure there is a complex interaction between the subunits that may be important in catalysis. This possibility is supported by the location of conserved elements in the sequence.

The dUTPase enzyme (EC 3.6.1.23) was obtained using an overproducing genetic construct[4]. Crystals were grown at room temperature in a mixture of polyethyleneglycol PEG8000, 50 μM $MgCl_2$, 450 μM pyrophosphate and succinate buffer at pH 4.2 (ref. 5). At this pH, activity measurements are difficult, but dissolved crystals show enzyme activity from pH 5. These crystals diffract X-rays to beyond 1.7 Å resolution. There is one subunit per asymmetric unit (data collection summarized in Table 3). Details of the crystallographic work will be published elsewhere.

dUTPase is packed as a trimer around the threefold axis of the crystal, with large solvent channels between the enzyme molecules. Chromatography confirms that the molecular weight of dUTPase in solution corresponds to a trimer (G.L., manuscript in preparation), rather than to a tetramer as reported earlier[6]. A trimeric subunit arrangement has also been found for a mammalian dUTPase[7]. The trimer has a wedge-shaped appearance when viewed perpendicularly to the 3-fold axis and a triangular face, typical for a trimer, when viewed along the 3-fold axis (Fig. 1*a*). The largest distance across a triangular face is about 60 Å and the length along the axis is ~45 Å. The subunits associate through interactions between twisted β-sheets, thus burying three hydrophobic surfaces about the 3-fold axis. The arrangement of monomers resembles that of the trimeric tumour necrosis factor[8,9], but the details of relative orientation of strands along the axis differ.

Figure 2*a* and *b* shows a topology diagram and a ribbon representation of the subunit which consists of a polypeptide chain of 152 amino acid residues; the first 136 of these are visible in the crystal structure. The remaining 16 residues up to the carboxy terminus contain several conserved side chains, but cannot be seen in the X-ray structure. The subunit is made up of about 40%, mainly antiparallel, β-pleated sheet and has a single short α-helix. The major tertiary structural feature is a sandwich made by two β-pleated sheets, 10–13 Å apart. The volume between them composes the hydrophobic core of the subunit. Two antiparallel β-strands, which are separate from the sandwich, together with one short helix and a long loop, complete the fold of the subunit. The structure resembles an 8-stranded jelly-roll composed of parts from two subunits and distorted by the direction of strand 1 (the amino terminus) being opposite to that of a regular jelly-roll[10].

The first two methionines in the dUTPase sequence, Met 1 and Met 2, are poorly defined in the electron density map. From residue 3 to 136 (Fig. 1*b*) the initial map was easy to interpret. Density is clearly visible for the side-chain ring of Phe 136, partly embedded among side chains of a symmetry-related subunit in the trimer (Fig. 3), but electron density abruptly stops after this residue. The remaining 16 amino acids of dUTPase are nevertheless present in the protein crystals (G.L., manuscript in preparation). Phe 136 lies on the surface of the trimer exposed to solvent channels in the crystal, and the 'missing' residues are presumed to be disordered.

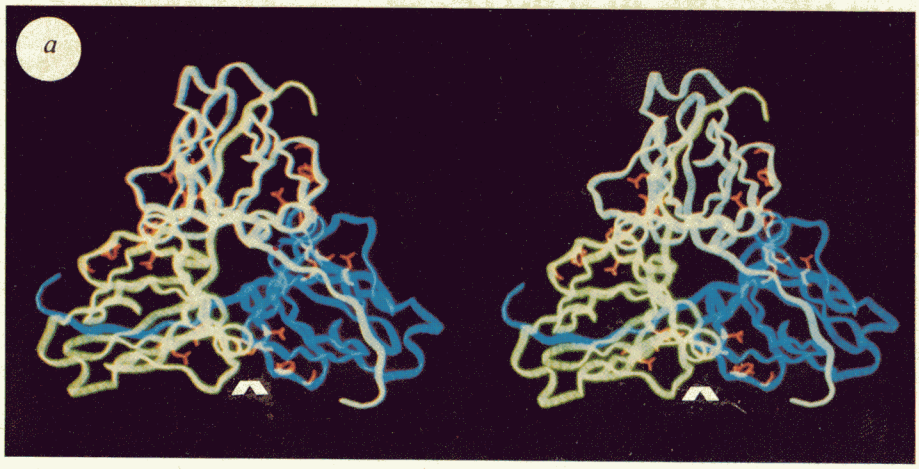
The interface between adjacent subunits in the trimer is made up by residues from a number of different parts of the sequence. (1) a hydrophobic interaction around the 3-fold axis at the centre of the trimer involves residues 47–58, β-strand 3 in the topology diagram, and 96–103, β-strand 6. Along the central axis, side chains of methionines, leucines, isoleucines, valines and threonines are closely packed. (2) The residues 20–30 of the loop region in one subunit form a contact to the bend 90–94 in another subunit (Fig. 2*b*). (3) A 35 Å-long carboxy terminal arm (residues 127–136) extends from the core of one subunit and makes extensive contacts along the surface of the neighbouring subunit (Fig. 3*a*). It constitutes strand 8 in the jelly-roll model (Fig. 2*a*), and forms the fifth strand in one of the β-sheets. A similar situation is encountered for the β-subunit of heavy riboflavin synthase[11], where the amino terminus forms the fifth strand in the β-sheet of a neighbouring subunit in a pentamer.

The enzyme requires divalent metal ions such as Mg^{2+} for activity, and the activation constant for Mg^{2+} ion and the inhibition constant for the cleavage product, pyrophosphate, have been reported for *E. coli* dUTPase[12]. These ions were included in the crystallization medium with the intention of forming a complex allowing the identification of the active site. Mg^{2+} and pyrophosphate cannot be seen in the electron density maps.

TABLE 1 X-ray data statistics

	Resolution (Å)	λ (Å)	*R*-merge (%)	Unique reflections	Metal sites per subunit
Native E(Mg-PP_i) data:					
2 crystals	1.9	1.009	6.9	13,640 96% complete	None found
Derivatives:					
Hg, 1 crystal	2.0	0.995	5.9	11,730	1
Pt, 1 crystal	2.1	1.050	5.7	9,910	1

The space group is *R*3 with cell dimensions $a = b = 86.6$ Å and $c = 62.3$ Å, and the crystals contain one subunit per asymmetric unit. Intensity data were collected using monochromatic synchrotron radiation at the EMBL outstation at DESY, Hamburg. The Hendrix–Lentfer imaging plate scanner, constructed in-house at EMBL, was used as detector. Data were processed with a modified version of the MOSFLM package. *R*-merge is defined as $\sum |I_i - \langle I \rangle| / \sum \langle I \rangle$, where I_i is an individual intensity measurement and $\langle I \rangle$ is the mean intensity for this reflection. Derivative data were collected at wavelengths chosen to maximize the anomalous signal. For the mercury derivative, a native crystal, E(Mg–PP_i), was soaked in 75 μM ethyl mercury phosphate for 4 h. The second platinum derivative was similarly prepared using 1.2 mM K_2PtCl_4 for 7 h. The mercury compound reacted with the only cysteine residue present in the sequence, whereas platinum binds to a methionine side chain. Both isomorphous and anomalous components were used for the phase determination to 2.2 Å resolution, giving an overall figure of merit of 51%, which was increased to 78% by solvent flattening[16]. This gave a high-quality map with clear continuity in the electron density. The model was built using the program FRODO[17] on an Evans and Sutherland PS330 graphics system. The initial model was refined by a cycle of simulated annealing using XPLOR[18], and subsequently by restrained least-squares minimization using PROLSQ[19]. The final *R* factor for 136 residues and 189 water molecules is 14.5% when all reflections between 1.9 and 8.0 Å are used. The geometry of the final model is in good agreement with the target values for the stereochemical parameters: the deviation in bond lengths was 0.010 Å and in torsion angles 2.5°. The average temperature factor for the protein atoms was 20.9 and for the solvent molecules 38.1.



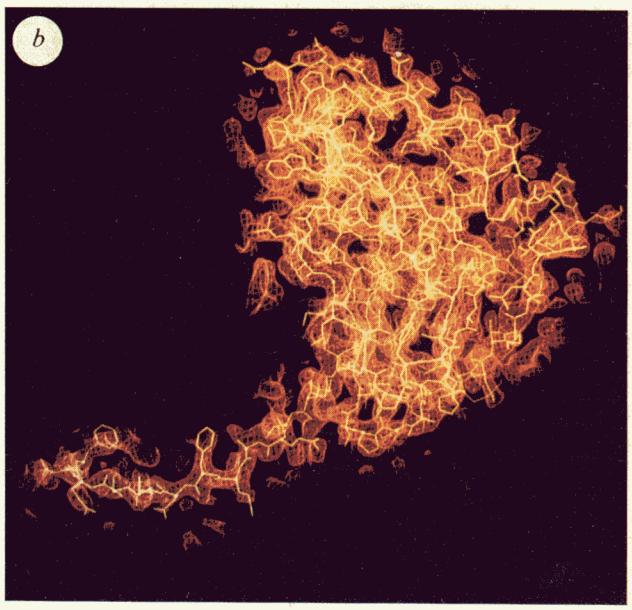
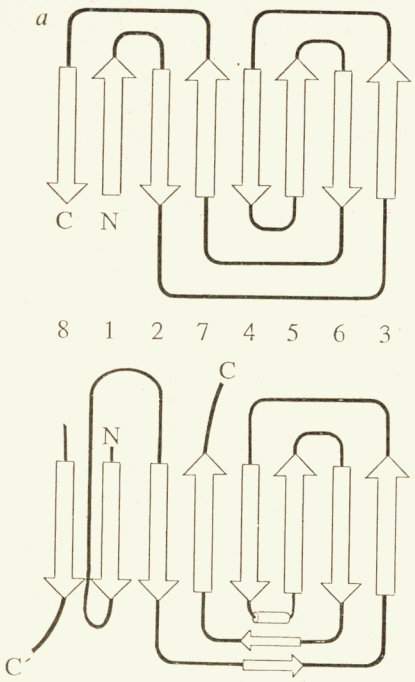


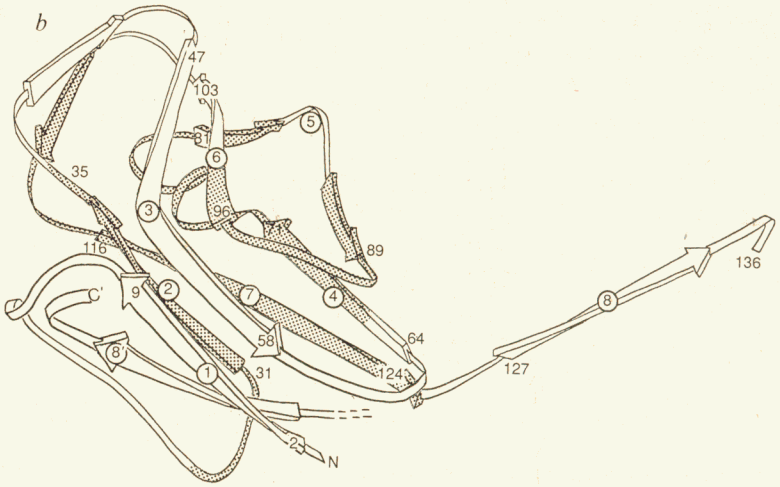
FIG. 1 *a*, Stereo view along the 3-fold axis of a ribbon diagram of trimeric dUTPase. The three subunits are shown in different colours. Side chains are only shown for those conserved residues that are accessible to solvent in the trimer. The depression on the surface proposed to contain one active site of dUTPase is marked by the white arrows. *b*, Electron-density map of the wedge-shaped subunit viewed at an angle to the 3-fold axis, with the model of 136 residues superimposed. Although hydrophobic residues pack closely within the jelly-roll, this core at the centre of the subunit is surrounded by small hydrophobic cavities. The synthesis is contoured at a level of 1 σ above the average density.

Given the high resolution and low *R*-factor (Table 1), this clearly shows that they are not bound in an ordered manner in our crystals, suggesting that the constants estimated differ from binding constants, at least in the crystal. This might be coupled to the fact that the carboxy terminus is flexible. Under certain conditions, the disordered residues may partake in the formation of a binding site for the nucleoside triphosphate that is made up from elements from more than one subunit.

Sequence data on dUTPases have mainly been derived from the DNA of the corresponding gene. The sequence of the *E. coli* enzyme[13] has been compared with that of three different herpesviruses[14] (herpes simplex, herpes zoster varicellosus, Epstein–Barr virus). Five regions, scattered along the sequence, are highly conserved (Fig. 2*b* and *c*), indicating that those parts are important for the catalytic function and/or stability. An inspection of the three-dimensional model of *E. coli* dUTPase reveals that some of the conserved residues are located in bends and loops that are in van der Waals contact. The sequence of residues 137–152 (not visible in the electron density maps) is particularly rich in glycines and shows significant homology not only to viral dUTPases but also to phosphate-binding sequences in other nucleotide-binding proteins[15]. But, the dUTPase subunit does not show the 'classical' Rossmann fold and thus represents a novel polypeptide fold for nucleotide binding.

Three shallow depressions are observed on the surface of the trimer. They can be seen at the interface between two of the subunits when viewed down the 3-fold axis as in Fig. 1*a*. The side chains of eight conserved amino acids (residues Gly 30, Asp 32, Ser 72, Ile 89, Asp 90, Tyr 93, Gly 95 and Gln 119), which have at least 10 Å^2 surface area accessible to solvent in the trimer, are seen to lie in and around these depressions. Moreover, the end of the extended arm (Phe 136) of the third subunit points towards this region (Fig. 3*b*). The conserved invisible carboxy terminus (Fig. 2*c*) can thus be imagined as interacting with the shallow cavity region. Although the active site cannot be identified with certainty, the information accumulated so far indicates that it could lie in that region. □





c

```
                10                  20
MMKKIDVKILDPRVGKEFPLPTYAT
 1
        30                  40                  50
SGSAGLDLRACLNDAVELAPGDTTL
  2
                60                  70
VPTGLAIHIADPSLAAMMLPRSGLG
  3                   4
            80                  90                  100
HKHGIVLGNLVGLIDSDYQGQLMIS
      5                   6
                110                 120
VWNRGQDSFTIQPGERIAQMIFVPV
  6                   7
                130                 140                 150
VQAEFNLVEDFDATDRGEGGFGHSG
  8
RQ
```

FIG. 2 *a*, Topology diagram of an idealized jelly-roll fold[10] with strand numbers indicated (top) compared with that of the dUTPase subunit. N and C, amino and carboxy termini; C′ indicates the carboxy terminus of an adjacent subunit. *b*, Schematic drawing of the subunit. Arrows represent β-strands. The first and last amino-acid residue of each strand within the jelly-roll barrel are labelled. Strand 8′ is from a neighbouring subunit. Shaded parts of the structure correspond to conserved sequences. *c*, The amino-acid sequence of dUTPase from *E. coli*. Residues 137–152 are not visible in the electron density map. To confirm the existence of these last 16 residues, peptides generated by tryptic hydrolysis of dissolved crystals were isolated and subjected to Edman degradation. Sequence alignments of dUTPases from *E. coli* and various viruses indicate the presence of homologous regions. Strongly conserved regions are boxed, and correspond to the shaded parts in *b*. Underlined sequences, labelled with bold numbers, show β-strand numbering according to the topology diagram in *a* and the same convention holds for the strand numbering in *b*.

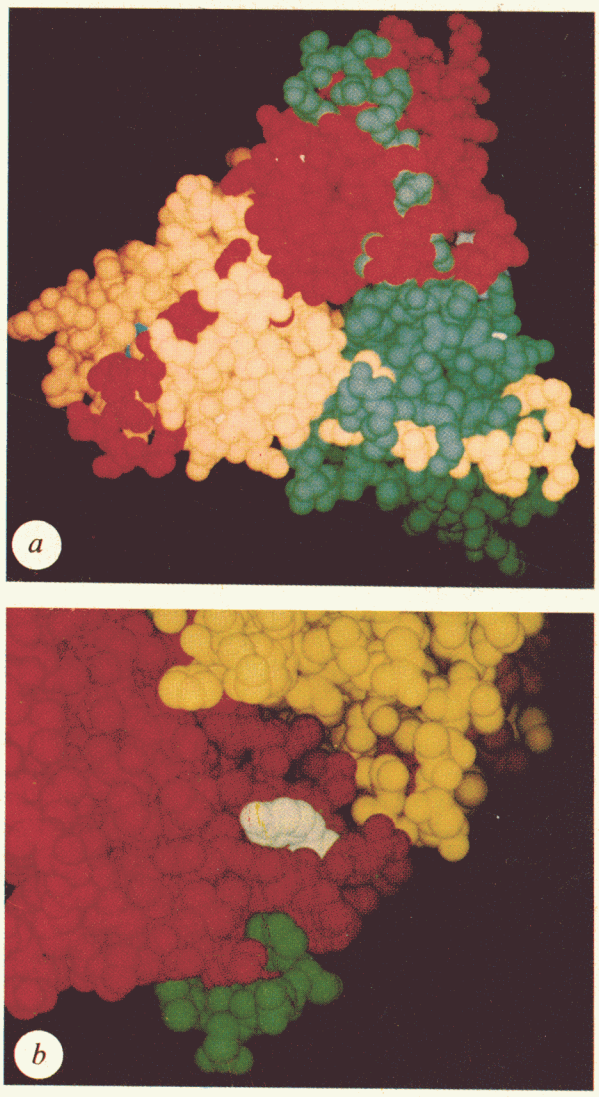


FIG. 3 Space-filling model of the dUTPase trimer viewed along the 3-fold axis with subunits colour-coded, showing carboxy terminal 'arms' running across the surface of a neighbouring subunit: for examples, a green 'arm' passing the red subunit and thereafter becoming disordered. *b*, The molecule rotated to show a close up of the intersubunit contacts at one of the three surface cavities. Evolutionarily conserved loops meet to form the depression comprising the bend 90–94 (red subunit, with the strictly conserved Tyr 93 coloured white) and the loop 20–30 (yellow subunit). Towards the depression a carboxy terminal 'arm' approaches (green subunit).

Received 5 September; accepted 19 November 1991.

1. Kornberg, A. & Baker, T. *DNA Replication* (Freeman, New York, 1991).
2. Tye, B. K. *et al. Proc. natn. Acad. Sci. U.S.A.* **74,** 154–157 (1987).
3. Bertani, L. E., Häggmark, A. & Reichard, P. J. *J. biol. Chem.* **238,** 3407–3413 (1963).
4. Hoffmann, I. *et al. Eur. J. Biochem.* **164,** 45–51 (1987).
5. Cedergren-Zeppezauer, E. S. *et al. Proteins: Struct. Funcn Genet.* **4,** 71–75 (1988).
6. Shlomai, J. & Kornberg, A. *J. biol. Chem.* **253,** 3305–3312 (1978).
7. Hokari, S. & Sakagishi, Y. *Arch. biochem. Biophys.* **253,** 350–356 (1987).
8. Jones, E. Y., Stuart, D. I. & Walker, N. P. C. *Nature* **338,** 225–228 (1989).
9. Eck, M. J. & Sprang, S. R. *J. biol. Chem.* **264,** 17595–17605 (1989).
10. Branden, C.-I. & Tooze, J. in *Introduction to Protein Structure* 70–74 (Garland, New York, 1991).
11. Ladenstein, R. *et al. J. molec. Biol.* **203,** 1045–1070 (1988).
12. Hoffmann, I. thesis, Univ. Saarbrücken, Germany (1988).
13. Lundberg, L. G., Thoresson, H.-O., Karlström, O. & Nyman, P. O. *EMBO J.* **2,** 967–971 (1983).
14. McGeoch, D. J. *Nucleic Acids Res.* **18,** 4105–4110 (1990).
15. Möller, W. & Amos, R. *FEBS Lett.* **186,** 1–7 (1985).
16. Wang, B. C. *Meth. Enzym.* **115,** 90–112 (1985).
17. Jones, T. A. *J. appl. Crystallogr.* **11,** 268–272 (1978).
18. Brunger, A. T. *J. molec. Biol.* **203,** 803–816 (1988).
19. Hendrickson, W. A. & Konnert, J. H. in *Biomolecular Structure. Conformation. Function and Evolution* Vol. 1 (ed. Srinivasan, R.) 43–57 (Pergamon, Oxford, 1980).

ACKNOWLEDGEMENTS. We thank A. M. Rosengren for enzyme purification facilities; C.-I. Bränden for suggestions and reading the manuscript; B. Furugren, S. Al-Karadaghi and T. Astlind for artwork; and M. Dauter for technical assistance. This work was supported by the Swedish Natural Science Research Council, Kungliga Fysiografiska Sällskapet i Lund, Swedish Cancer Society, and Wallenbergstiftelsens Jubileumsfond (G.L. and E.C.-Z.).